440-TP-007-001

Production Platform Families for the ECS Project

Technical Paper

May 1995

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Harvey Privor /s/

Harvey Privor, PDPS H/W Systems Engineer Date
EOSDIS Core System Project

SUBMITTED BY

Parag Ambardekar /s/

Parag Ambardekar, PDPS Manager Date
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Contents

Contents

1. Introduction

2. Executive Summary

3. Background

4. Design/Trade Alternatives

5. Analysis and Key Trades/Rationale

6. Selection Criteria and Evaluation

7. Recommendation/Conclusion

Figures

Tables

Abbreviations and Acronyms

This page intentionally left blank.

1. Introduction

1.1 Purpose

The purpose of this trade study is to recommend one or more processor classes for the ECS Data Processing Subsystem based upon primarily cost/performance criteria, and the ability to run parallel science software.

1.2 Scope

This study makes recommendations which focus on IR-1 and Release A with a "Look Ahead" to Release B for scalability and evolvability. The recommendations resulting from this study will be the basis for the Data Processing Subsystem hardware procurement process beginning in the latter part of the 1st quarter, 1995. The overall ECS hardware design philosophy is to support a heterogeneous computing environment.

The scope of this trade study is to select one or more class(es) of processors from a major family of computer platforms including:

- Uniprocessor Workstation/ Servers
- Workstation Farms
- Vector Supercomputers
- Parallel Processors

The parallel processor class was extended to include the following:

- Low-end Symmetric Multi-processors (SMP), or SMP-L. These are processors which have the capability to expand (up to 8), but are initially configured with a single CPU;
- High-end SMPs, or SMP-H. These are processors which have the capability to expand (usually to 12 or more CPUs) and are initially configured with a minimum of two CPUs;
- Massively Parallel Processors (MPP) that contain more than 64 processors
- SMP Clusters. This group consists of more than one fully configured SMPs.

Science Processing platform class(es) recommendations for ECS are primarily based upon: (1) Cost/performance tradeoffs, (2) Analysis of PDR Technical Baseline containing Ad Hoc Working Group on Production (AHWGP) data, (3) ECS Science and Technology Lab Prototyping, (4) Trade Studies , and (5) Scalability of hardware and evolvability of hardware and software, The candidate hardware is tailored to DAAC unique instrument processing needs. Suitability of candidate processor platform's ability for parallel software development and execution for future releases is a special criterion of particular importance that is factored into the recommendation. Algorithm development, test and maintenance costs and architectural design impacts are significant factors in the platform class(es) selection.

Analysis consists of evaluating the candidate platforms based on critical selection criteria (e.g. cost/ performance, scalability, risk, etc.) and quantifying the results. This study is conducted in two phases: Static analysis and dynamic analysis using the ECS System Performance Model. The first phase based on static analysis has been completed. The static model (spreadsheet analysis) derives from AHWGP data time-average processing (MFLOPs), I/O bandwidth, disk bandwidth, and disk storage. Upon completion of the validation of system modeling (system level dynamic model using AHWGP and other data as input) this study will be readdressed on an as needed basis to meet Release A Critical Design Review (CDR) and Release B goals as a minimum.

A variety of representative hardware platform processor characteristics have been determined. These characteristics have been gathered from the various vendors through a questionnaire/ survey.

Experience and lessons learned from the ECS STL Prototyping of Science Software (Ref: 194-00569 TPW, 430-TP-006-001) evaluations under various processing paradigms provided inputs for identifying production platform families for Data Processing Subsystem hardware selection. The Science Software Execution Prototype used science software to study the various processing alternatives (e.g. DCE, SMP, DMP/Workstation cluster, and MPP).

Key trade studies related to the platform class(es) recommendations (Ref. 211-CD-001-002) include: (1) Distributed and Parallel Processing (Ref: 440-TP-008-001), (2) Production Topologies (Ref: 440-TP-006-001), and (3) Production Platform Families. The Distributed and Parallel Processing trade study investigates various processing alternatives to ECS science algorithms, examining the benefits of distributed and parallel computing. Production Topologies analyzes physical processing topologies that can impact hardware requirements, overall performance, network capacity, throughput, and staging storage. Results from the Production Topology study will be used to support Release B hardware selection.

1.3 Organization

This document consists of eight sections. Section 1 contains the introduction consisting of purpose and scope, background and document organization, review and approval, and applicable documentation. This section also contains a brief description of the trade alternatives and associated issues. Section 2 is the executive summary providing a high level summary of sections 3 through 7 including background, design/trade alternatives, analysis and key trades, selection criteria and evaluation, and recommendations/ conclusions. Section 3 briefly describes the Data Processing Subsystem hardware configuration items. Section 4 contains the design/trade alternatives of processor classes including workstations/ servers, vector supercomputers, and parallel processors. Section 5 discusses the analysis and the inputs from the modeling and prototyping efforts. It also discusses the application of the technical baseline including the Ad Hoc Working Group for Processing (AHWGP) inputs and phasing. Section 6 identifies the selection criteria (e.g. cost, performance, risk, RMA, scalability, evolvability and suitability) and associated weighting factors. A Figure of Merit (FOM) evaluation is described quantifying the weighting and scoring for each criteria. Section 7 provides the recommendations and conclusion for the selection of the processor class. Section 8 includes definition of acronyms.

1.4 Acknowledgments

The author gratefully acknowledges the excellent support received from Electronic Data Systems, specifically from Barbara Schroeder and Chuck Pross, for their ideas, knowledge and assistance.

1.5 Review and Approval

This trade study is an informal document approved at the Office Manager level. It does not require formal Government review or approval. However, it is submitted with the intent that review and comment be forthcoming. The intent of this document is to support the SDPS segment design specification, support the procurement process, and provide definition of the Data Processing Subsystem processor class for PDR. Questions regarding technical information contained within this trade study should be addressed to the following ECS and GSFC contacts:

ECS Contacts: Harvey Privor

(301) 925-0578

hprivor@eos.hitc.com

GSFC Contacts: Tonjua Hines, PDPS H/W Manager

(301) 286-8807

thines@ulabsgi.gsfc.nasa.gov

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616 McCormick Drive
Landover, MD 20785

1.6 Applicable and Reference Documents

1. Technical Baseline, November 8, 1994; J. Guzek, CCR No. 94-0132
2. PDPS Prototyping at ECS Science and Technology Laboratory; Progress Report #4, Technical Paper 194-00569TPW, September, 1004, Narayan Prasad
3. Production Topologies: A Trade Analysis, Technical Paper (Draft), March 14, 1995, Narayan Prasad, PDPS Scientist Engineer
4. Science and Data Processing Segment (SDPS) Requirements Specification for the ECS Project, 304-CD-002-002, March 1995
5. Science Data Processing Segment (SDPS) Segment Design Specification, 305-CD-002-002, March 1995
6. Trade-off Studies Analysis Data for the ECS Project, 211-CD-001-002

7. Portability Issues for ECS Science Software Integration, Narayan Prasad, Science Software Integration and Test Workshop, 726-PP-002-001, April 18, 1995
8. PDPS Prototyping at the ECS Science and Technology Laboratory: Report #5, Technical Paper, April 1995, Narayan Prasad, 430-TP-006-001
9. Systems Performance Model for the ECS Project, 241-TP-001-001
10. PDPS Prototyping at the ECS Science and Technology Laboratory: Report # 5, April 1995, 430-TP-006-001

2. Executive Summary

Data Processing Subsystem platform classes are selected for each DAAC site by release. A primary focus is on Ir1 and Release A with a "Look Ahead" to Release B for scalability and evolvability. The processor classes are selected based on primary evaluation criteria of cost/performance, risk, scalability, and evolvability. Suitability of the candidate processor platform (how it operates with parallelized science software and phasing of performance requirements) is a special criterion of particular importance that has also been factored into the recommendation.

The major design drivers are as follows: (1) Cost/performance, (2) Processing, I/O bandwidth, disk storage derived from static analysis of AHWGP data, (3) Scalability and evolvability to Release B and beyond, (4) Ease of algorithm development, integration, and maintenance, and (5) Separation of AI&T environment from operations environment.

The platforms which suit the requirements for Ir1, Release A and the initial phases of Release B are the high-end SMP (SMP-H) for LaRC, EDC, and GSFC and the Uniprocessor Workstation/Server for MSFC (see Table 2.1-1). The SMP is proposed because it offers good price/performance, handles parallelization of science software, and is scalable and evolvable. Since LaRC and MSFC are operational sites during Release A, a second processor is provided to support a fail soft backup requirement. For Release A sizing, Science Processing platform class(es) recommendation for ECS is primarily based upon PDR Technical Baseline containing Ad Hoc Working Group on Production (AHWGP) capacities (November 1994), scalability of hardware, evolvability of hardware and software, and cost/performance. Processor platform class(es) selection will be readdressed on an as needed basis to meet Release A CDR and Release B goals at a minimum. The processing requirements for each DAAC were examined (peak MFLOPs derived from the technical baseline, I/O bandwidth derived from static analysis and disk volume derived from static analysis) and the following recommendation is presented. DAAC unique characteristics of the processor platforms are provided in the SDPS Segment Design Specification (Ref: 305-CD-002-002), Sections A.1-A.3 in the DAAC Unique Appendices for the operational sites (e.g. GSFC, MSFC, and LaRC).

The Algorithm Integration and Test Hardware (AITHW) supports DAAC operations users performing (and/or assisting instrument team members with) science software integration, systems validation, and integration and test. It is important to note that the workstation-based operations are performed by the AITHW, while the prime science software integration and test capacity are provided by the science processing hardware (SPRHW). In summary, the AITHW supports the DAAC configuration, control, and management of the AI&T processes engaged on the target science processors. An additional processor platform is provided at LaRC and MSFC for backup purposes in Release A.

The Data Processing Subsystem hardware recommendation is derived from static analysis of AHWGP data and preliminary validation results from the System Performance Model (Ref: 241-TP-001-001) for Release A. The AHWGP data incorporates ESDIS phasing factors. This recommendation encompasses results from the ECS Science and Technology Lab Prototyping and the previously mentioned trade studies.

Table 2.1-1. Platform Recommendation for Ir1 and Release A

Site	Platform Class	Supporting
LaRC	SMP-H + SMP-H + Uniprocessor workstation (Queuing)	Ir1/Rel A Rel A
MSFC	Uniprocessor workstation +Uniprocessor workstation +Uniprocessor workstation (Queuing)	Ir1/Rel A Rel A
EDC	SMP-H No Change in Science Processing +Uniprocessor workstation (Queuing)	Ir1/Rel A Rel A
GSFC	SMP-H +CPUs +Uniprocessor workstation (Queuing)	Ir1/Rel A Rel A

The cost/performance tradeoffs are a major factor in the selection of the processor platform classes.

Vendor averaged $/ MFLOPs for a high end SMP is $176/ MFLOPs, which is less than one- half the cost./ performance of a vector supercomputer. Similarly, the cost performance of the SMP compares very favorably to the workstation class. SMPs provide a cost effective solution that is low risk with a single operating system and relative ease of programming with capability to process sequential science algorithms and transition to parallel processing of these algorithms.

The system design will support a phased procurement, heterogeneous architecture, and multivendor platforms.

As the processing requirements become more mature and dynamic system modeling results become available, the analysis supporting the recommendation will become more accurate. Platform class(es) selections are not expected to change for Releases IR-1 and A, but the number of processors recommended, the amount of memory and I/O bandwidth estimated, or the size of disk are the components most likely to be refined. It is anticipated that the platform processor class(es) will be re-evaluated for Release B and subsequent releases.

3. Background

The Data Processing Subsystem, in conjunction with the Planning Subsystem, plans for and allocates resources to any task suited to the inventory of available resources (e.g. on demand) with Planning providing production planning and production management. The Data Processing Subsystem performs data production and provides the capabilities to maintain and control the processing. Processing data production functionality includes: (1) provides input data for science software access, (2) executes science software, and (3) provides output data for data server storage. The Data Processing Subsystem, in addition to its 4 CSCIs, is composed of three HWCIs: Science Processing, Algorithm Integration & Test (AI&T), and Quality Assessment (QA) Monitoring.

The Science Processing HWCI supports the managing, queuing and execution of processes at each DAAC site. It contains staging (working storage), input/output (I/O), and processing resources necessary to perform routine processing and subsequent reprocessing. The Science Processing HWCI consists of two components: (1) Science Processing, and (2) Processing Queue Management. The Science Processing component is broken up into processing "strings", which are chains of processing, I/O, and staging resources configured to deal with unique processing requirements. A processing string may be assigned to one specific class of instruments alone. This HWCI contains the processors that support science processing, as well as algorithm integration & Test and QA. During Ir1, the processor will only support algorithm integration & test and systems integration & test. In Release A, one processor will be dedicated to science processing while a second processor supports AI&T in order to separate the AI&T environment from the operations environment. In the event of a science processor failure, the AI&T processor will be used as a backup.

The Algorithm Integration & Test (AI&T) HWCI provides hardware resources to support DAAC operations and users performing science data algorithm integration and test (AI&T), systems validation and systems integration and test (I&T). It should be noted that the AI&T HWCI provides the workstation based operations support hardware, while the prime science software integration and test capacity is provided within the Science Processing HWCI. The AI&T HWCI only provides the operations support workstations to permit DAAC personnel to configure, control, and manage the AI&T processes engaged on the target science processors. ESDIS phasing factors estimate a 0.3 X, for pre-launch AI&T starting 2 years before launch, where X is defined as at - launch processing.

The Quality Assurance (QA) HWCI contains hardware resources to support DAAC operations and users performing planned and routine quality assurance of product data. QA processing requirements are currently being jointly evaluated with the investigative teams. Current operational assumptions include DAAC QA process being performed at the sites in conjunction with SCF-based QA. The current QA HWCI design baseline includes QA monitors.

4. Design/Trade Alternatives

Candidate processor classes evaluated for Ir1 and Release A are:

- Uniprocessor workstations/servers
- Workstation farms/clusters
- Vector Supercomputers
- Parallel Processors

The parallel processor class was extended to include the following:

- Low-end Symmetric Multi-processors (SMP), or SMP-L. These are processors which have the capability to expand (up to 8), but are initially configured with a single CPU;
- High-end SMPs, or SMP-H. These are processors which have the capability to expand (usually to 12 or more CPUs) and are initially configured with a minimum of two CPUs;
- Massively Parallel Processors (MPP) that contain more than 64 processors.
- SMP Clusters. This group consists of more than one fully configured (or fully populated) SMP-H.

4.1 Uniprocessor Workstation/Servers

The uniprocessor workstation/server class is a low cost commodity item configured with a single CPU (either CISC or RISC). By itself, it has limited scalability. They are flexible, often under individual user control and provide the user with a dedicated resource. They can be used alone or in clusters. Examples include, but are not limited to the following: SGI Indigo R4000, DEC Alpha AXP varieties, IBM RISC/6000 varieties, Sun SPARC system varieties, and Hewlett Packard 9000 7xx varieties. It should be noted that the industry is moving away from uniprocessors. The cost of adding processors is making the multiple processor "box" more appealing.

4.2 Workstation Farms

A workstation farm is a cluster of workstations connected via a LAN often providing excellent scalability. However, low network bandwidth and high latency can be limitations. SMP clusters connected by a high performance switch can be considered as an extension of this class. Utilization of a High Performance Parallel Interface (HIPPI) switch or Fiber Channel can significantly enhance network interconnect bandwidths. This connectivity provides excellent scalability.

4.3 Vector Supercomputers

These are supercomputers with the vector processing instruction set, real physical shared global memory tuned to floating point processing where peak MFLOPs are greater than peak MIPS. The vector supercomputer allows a single instruction to do a complex vector computation. Examples include, but are not limited to the following: Cray C90, Cray YMP, Convex 3800, and Fujitsu vp2600. Vector supercomputer cost and suitability do not match the processing requirements for Ir1 and Release A. However, this category is a potential platform candidate for Release B and will be re-evaluated at that time.

4.4 Parallel Processors

The parallel processor class is a computer consisting of two or more commodity based CPUs (either CISC or RISC) which can function either as a workstation or a server. A further breakdown of this class provides the massively parallel processors (MPPs) and symmetric multiprocessors (SMPs) as potential platforms. Examples include, but are not limited to the following: Cray T3D, Convex Exemplars, SGI Power Challenge XL, SGI Challenge, DEC Alpha AXP, IBM SP2, DEC 2100, DEC TURBOlaser.

Symmetric Multiprocessors (SMPs)

Symmetric Multiprocessors (SMPs) provide excellent cost/performance, graceful degradation, good scalability, and load balancing. This class is selected on the basis of a cost effective solution that meets both performance and scalability requirements. The SMP can be viewed as "multiple workstations in a box" with all processors identical, and providing coherent shared memory. Interconnect bandwidths within the SMP are very high. The SMP can process sequential science algorithms with the capability to transition to parallel processing of these algorithms.

Massively Parallel Processors (MPPs)

Massively Parallel Processors (MPPs) contain numerous processors (usually more than 64 processors). MPPs provide high speed interconnects and can provide multiple I/O channels. MPPs require significant programming effort to tune this platform to perform at peak efficiency for parallel algorithms.

5. Analysis and Key Trades/Rationale

5.1 Introduction

Hardware recommendations are based upon: (1) Cost/performance tradeoffs, (2) Analysis of AHWGP data incorporating ESDIS phasing and efficiency factors, (3) Prototyping studies, and (4) Trade studies. The hardware selection and the design supports a phased procurement, heterogeneous architectures, use of heritage software, and multivendor platforms. Additional criteria for selecting hardware platform classes include technical and schedule risk, reliability/maintainability/availability, portability, and suitability.

This section provides the hardware configuration item (HWCI) design rationale for the proposed high end SMPs (SMP-H) for LaRC, EDC, and GSFC and the uniprocessor workstations for MSFC for Ir1 and Release A. The SMP supports sequential processing, in addition to symmetric and distributed memory parallel processing paradigms. This platform is analogous to "multiple workstations in a box", providing a coherent shared memory and containing identical processors with a single operating system. The cost/performance tradeoff (see section 5.3.1), shows SMPs provide the best price/performance with more capacity per box for the money. SMPs also have very good scalability, relatively high I/O bandwidth (e.g. 1.3 GB/s Peak), and very high interconnect bandwidths.

5.2 Assumptions

The following assumptions have been made in this trade study:

- The November 1994 Technical Baseline is used
- Processing MFLOP requirements have been "phased" and have been increased by factor of 4 to account for machine cycles inefficiencies
- The uniprocessor class is assumed to consist of workstation/server equipment classes
- Ir1 support includes:
    - interface testing and the initial phases of Algorithm Integration and Test (AIT) of instrument science code for Release A
    - AIT support for ASTER (EDC) on the AM-1 platform (scheduled for Rel B)
    - AIT support for MODIS (GSFC) on the AM-1 platform (schedule for Rel B)
- Releases can be correlated to the following dates:
    - Ir1 Jan 95 - Dec 95
    - Rel A Jan 96 - Dec 96
    - Rel B Jan 97 - Dec 97
- No Quality Assurance monitoring is included in these processing capacities

5.3 Analysis and Key Trades

The SPRHW design analysis for PDR was accomplished through a number of distinct but highly related efforts. These efforts included: cost/performance trade-off analysis, key design trade studies, prototyping, and joint analysis of requirements with the AHWGP.

5.3.1 Cost/Performance Tradeoff

A cost/performance trade-off has been performed for the various candidate processor platform classes. The analysis is based on current technologies. Costs are list prices in 1995 dollars ($). Representative vendor average list prices are normalized to peak MFLOPs performance for each of the platform classes. The SMPs are broken down into two categories: SMP-L (1 to 8 CPUs) and SMP-H (2 to 64 CPUs). The main distinction between the high- and low-end SMPs are processing power, scalability, and cost. SMPs are further classified as minimum and maximum configurations, where minimum refers to the least number of CPUs configured for the SMP (e.g. 1 CPU for SMP-L, and 2 CPUs for SMP-H) and maximum refers to a fully populated SMP. The $$/MFLOPs is taken as an average normalized cost of the minimum and maximum SMP configurations.

Considering the AHWGP required processing performance capacities as a constraint, the most cost-effective processor platform selection for the LaRC, and GSFC DAAC sites is the high-end SMP (SMP-H) with an average price per performance of $176 per MFLOP as shown in Table 5.3-1. Although, the low-end SMP cost/performance ratio is less than the high-end, this class of SMP cannot satisfy the MFLOPs nor the scalability for LaRC and GSFC (see Table 5.3-2). The SMP-H is approximately 23% and 60% of the normalized cost (price/performance) of the Vector Supercomputer and the MPP, respectively. Comparison of price/performance is actually more favorable for the Release A maximum configuration SMP-H (i.e. $134/ MFLOPs).

Table 5.3.-1 Price/Performance

Platform Class	Peak MFLOPs	Average $$/MFLOPs
Uniprocessor Workstation	100	205
SMP-L		146
Min.	250	
Max.	875	
SMP-H		176
Min.	850	
Max.	6,000	
Vector Supercomputer	1,600	751
MPP	25,000	300

The AHWGP required MFLOPs capacities by release and DAAC site are summarized in Table 5.3-2, which provides insight into machine sizing requirements to support the cost performance tradeoff. ESDIS phasing and machine efficiency factors have been applied to these numbers. Processing requirements for EDC can be satisfied by a low end SMP (SMP-L) for IR-1 and Release A. However, it can be shown that to achieve the scalability for Release B, it is more cost effective to procure a SMP-H. The MSFC processing requirements can easily be satisfied by an uniprocessor workstation class. The average list price of the workstations surveyed is less than 1/3 the cost of a minimum configuration SMP-L.

Table 5.3-2 AHWGP Required Processing Performance Capacities

DAAC	Release	MFLOPs (Peak)	Activity
LaRC	Ir1 Rel A	1100 + 2333	AI&T Science Processing operations
MSFC	Ir1 Rel A	10 No Change	AI&T Science Processing operations
EDC	Ir1 Rel A	120 + 57	AI&T AI&T
GSFC	Ir1 Rel A	2150 + 2040	AI&T AI&T

This sizing effort was accomplished in conjunction with the performance requirements provided within the PDR Technical Baseline (covering Ir1 and Release A). A complete discussion of these requirements is provided within the SDPS Requirements Specification for the ECS project (Ref: 304-CD-002-001).

The primary cost criteria used in this analysis is $$/MFLOPs, which is a normalized cost based on 1995 list prices and current technology.

Development effort is not in terms of dollars, but in terms of programming difficulty. For example, if a highly parallelized environment is required, and compilers are not robust enough to handle the parallelization effort, the burden is on the developers.

Life cycle cost is a combination of several factors:

- acquisition cost;
- maintenance cost; and
- support cost.

Acquisition cost is a rough estimate of the initial equipment cost normalized as a function of processing power, cost/millions of floating point operations per second (MFLOPs). List price is used as the basis.

Maintenance cost is estimated for the equipment over the life of the contract. This is an approximation of maintenance based on the program's current hardware maintenance contract.

Support cost is the estimated operational cost of the equipment over the life of the contract.

5.3.2 Trade Studies

The "Trade-off Studies Analysis Data for the ECS Project" document (Ref: 211-CD-001-001) provides an overview of the related trade studies which are briefly described below.

A *Distributed and Parallel Processing Trade Analysis* was performed examining the benefits of distributed and parallel computing. The trade studies various processing alternatives for ECS science algorithms and provides up-to-date information on processing technologies. This trade analyzes the applicability of using OSF/ Distributed Computing Environment (DCE), SMP, DMP (including workstation cluster), and MPP for ECS science software.

A *Production Topologies Trade Analysis* examines the advantages and disadvantages of distributing processing tasks from one or more instruments across one or more processing strings. The resulting recommendation will provide a cost effective way of distributing processing to maximize throughput, minimize data movement, and provide and retain the flexibility to evolve with changing processing requirements. This trade analyzes physical (not logical) processing topologies that can impact hardware requirements, overall performance, network capacity, throughput, and staging storage. A key recommendation from this trade study is that multiple strings/cluster/subnetwork formation alternatives be allowed between DAAC sites and within them. Recommendations for hardware selection based on cluster optimization alternatives for Release B and beyond are made.

A *Production Platform Families Design Trade* study has been performed to recommend one or more Science Processing HWCI platform processor class(es) based on the Technical Baseline/AHWGP data, scalability, risk, and cost. This trade provides the basis and rationale for Science processing HWCI hardware class recommendation for IR-1 and Release A with projection to Release B. The recommendations resulting from this study are the basis for the Data Processing procurement process scheduled for early 1995.

Three major considerations are examined along with quantitative selection criteria to come up with the final platform recommendations.

- Phased performance requirements
- Algorithm development, test and maintenance costs
- Architecture design impacts

This trade study is being conducted in two phases: Static analysis (spreadsheet analysis) and system modeling. The result of both phases is data which is used in sizing the processing platforms. The first phase (based on static analysis) has been completed and is summarized here. Upon completion of the validation of system modeling, this study will incorporate the dynamic results of this model and be re-issued.

To meet the needs of this study for Release A PDR, static modeling results were derived using data which is a combination of processing efficiency factors and processing phasing requirements applied to the Technical Baseline of November 1994. The AHWGP data provided by the instrument teams has been incorporated into this Technical Baseline as well.

Analysis of January 1995 Technical Baseline information, including AHWGP data, is under way. In addition, further refinements to the baseline through the AHWGP (especially for areas including: QA, reprocessing, and Release B sizing changes) will result in repeated study analysis throughout the the Release A CDR phase (planned) and beyond. The analysis will be further refined based on dynamic model results. Further detail regarding AHWGP data as applied to this design is discussed in Section 5.3.4.4.

In addition to the MFLOP processing requirements, the I/O bandwidth and disk volume capacities were calculated using the results from static modeling and are "time averaged" for epochs c (IR1/Release A time frames) and k (Release B/C time frames). Data volumes for a 24-hour period were averaged into MB/second. The peak disk volume was derived from the AHWGP data by multiplying the volume at initiation by a factor of 2 (2 days worth of staging capacity).

This information was then analyzed (for each DAAC and instrument) and the following platform classes are recommended: SMP configurations for LaRC, EDC and GSFC, and uniprocessor workstation/server configurations for MSFC. The EDC and GSFC are provided hardware to support AI&T activities for IR-1 and Release A. The SMP solution is viable for IR-1 and Release A (and potentially Release B) because it offers a very flexible platform in which applications can run in either a serial or parallel mode and the SMP, by definition, is a reasonably scalable system. It also provides fast channel communications and is easy to administer.

5.3.3 Prototype Studies

Prototyping representative science algorithms at the ECS Science and Technology Laboratory (STL) using processing alternatives provided input and rationale for the selection of platform classes for the Science Processing HWCI. The Science Software Execution Prototype used science algorithms (e.g. Pathfinder, AVHRR/Land, SSM/I, SeaWinds) to study applicability of various processing alternatives (DCE, SMP, DMP/Workstation cluster, and MPP). The features of this prototyping effort incorporated distributed computing of Pathfinder AVHRR/Land using OSF/DCE on physically distributed workstation cluster. Multiprocessing, using SMP, DMP/ workstation cluster, of SSM/I and SeaWinds using automatic parallelization tools were demonstrated. SDP Toolkit performance studies were also conducted. The Science Software Execution Prototype also provided inputs for science software portability issues (e.g. 32 bits vs 64 bit architectures). New processing technologies were explored including architectures and software tools. The science processing prototyping activities provided hands-on experience with these newer technologies and also the rationale for recommending the most appropriate hardware for the DAACs.

5.3.4 Technical Baseline/AHWGP Analysis

This section describes the analysis of the Technical Baseline/AHWGP (see TP # 210-TP-001-002), which supports the evaluation of the various hardware platform classes and provides more rationale for the recommendation.

5.3.4.1 AHWGP Requirements Summary

Processing requirements as discussed in this Release A PDR effective volume, are based on Technical Baseline/AHWGP data of November, 1994. Further revisions of AHWGP data, provided and/or planned, will result in more analysis before CDR, and for Release B IDR. This data is provided in the SDPS Requirements Specification, (Ref: 304-CD-002-001). A roll-up summary table of the AHWGP requirements for the operational sites (i.e. LaRC and MSFC) for Release A is provided in Table 5.3-3. These rolled-up AHWGP requirements are based on raw data without ESDIS phasing factors nor machine efficiencies applied.

Table 5.3-3 Summary of AHWGP Requirements for Release A Operational Sites (Part 1 of 2)

Process	DAAC	Volume at Initiation (MB)	Staging I/O (MB)	Volume at Completion (MB)	Destaging I/O (MB)	I/O Req'ments (MB)	CPU Req'ments (MFPOs)	No. Input Files	No. Output Files	Activations (day-1)
CERES 10aT	LaRC	10,317	10,367	10,931	564	10,931	245,700	1738	1	0.03
CERES 11a	LaRC	91	91	182	0	182	37,800	1	1	0.1
CERES 12aF	LaRC	82	32	334	252	334	37,800	10	24	1
CERES 1aT	LaRC	138	87	852	714	852	20,790	4	25	1
CERES 2aT	LaRC	375	324	706	331	596	3,780	4	2	0.6
CERES 3aT	LaRC	164	114	836	672	836	47,250	3	3	0.03
CERES 4aF	LaRC	348	206	593	245	505	34,020	8	2	24.8
CERES 9aTF	LaRC	205	154	207	2	207	4,914	3	1	24.8
LIS	MSFC	7	7	93	86	92	2,492	2	11	14.56

Table 5.3-3 Summary of AHWGP Requirements for Release A Operational Sites (Part 2 of 2)

Process	DAAC	Volume Staged (MB/Day)	Volume Destaged (MB/day)	CPU req'ts MFPOs per day)
CERES 10aT	LaRC	311	17	7,371
CERES 11a	LaRC	9	0	3,780
CERES 12aF	LaRC	32	252	37,800
CERES 1aT	LaRC	87	714	20,790
CERES 2aT	LaRC	194	199	2,268
CERES 3aT	LaRC	3	20	1,418
CERES 4aF	LaRC	5,109	6,076	843,696
CERES 9aTF	LaRC	3,819	50	121,867
LIS	MSFC	102	1,252	36,284

5.3.4.2 Phasing

Application of the phasing factors to the AHWGP processing requirements is a very important step in each DAACs platform class evaluation. Scalability becomes a heavily weighted criteria when the processing at the DAAC ranges from low to high as science software is tested and integrated, instruments are launched and calibrated, and standard production and reprocessing begins. An example of how phasing is applied for CERES processing at LaRC and how the results enter into the platform recommendation is described in the next section, and an explanation of phasing factors as applied to standard products follows.

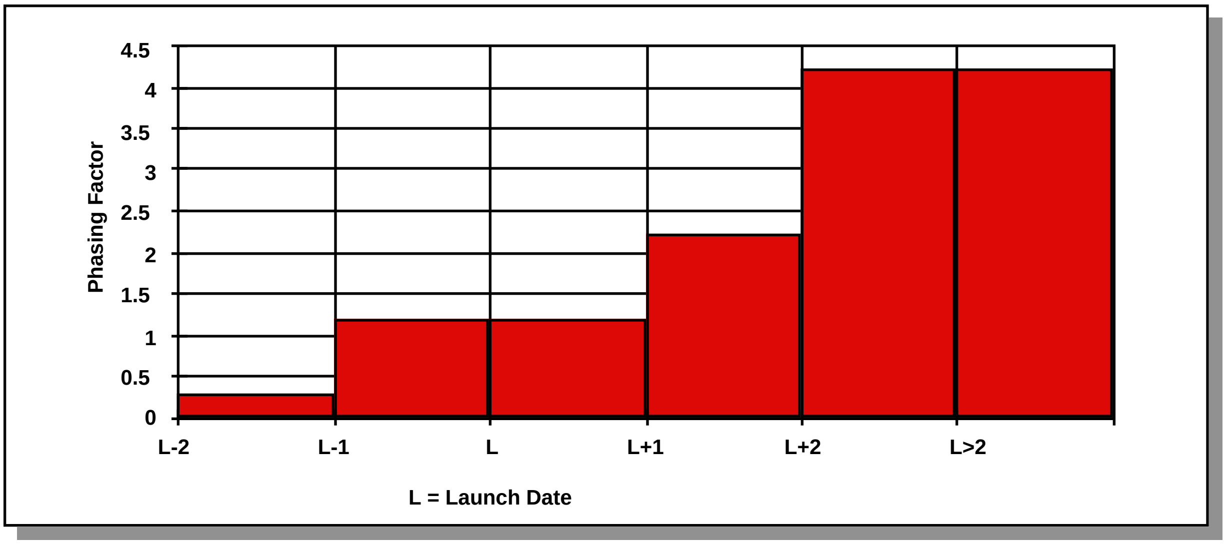

Figure 5.3-1 ESDIS Phasing Factors

Figure 5.3-1 summarizes and illustrates how phasing factors are applied to processing capacity based on launch dates. This phasing is applied to each platform and its scheduled launch date and incorporated into the Release schedule.

- **0.3X for L-2 < t < L-1** For pre-launch AI&T starting at launch minus 2 years, AI&T requires 0.3 of the processing estimate at launch during the period 1 to 2 years prior to launch. X is defined as at-launch processing estimate for pre-launch AI&T.
- **1.2X for L-1 < t < L+1** For pre-launch AI&T and system I&T, starting at launch minus 1 year, AI&T and system I&T requires 1.2 times the processing estimate at launch during the year prior to launch. Standard instrument processing requirements begin from launch date and last for the remainder of the life of the instrument. X is defined as at-launch processing estimate for prelaunch AI&T and systems I&T.
- **2.2X for L+1 < t < L+2** For post-launch AIT, standard processing, and reprocessing of data, starting at launch plus 1 year, 2.2 X is required. X is defined as the standard processing estimate for that period. The breakdown for this period is approximately 0.2 X for AI&T, 1 X for standard processing, and 1 X for reprocessing.

- **4.2X for t > L+2** For post-launch AI&T, standard processing and reprocessing of data, starting at launch plus 2 years, 4.2 X is required. X is defined as the standard processing estimate for that period. The breakdown for this period is approximately 0.2 X for AI&T, 1 X for standard processing, and 2 X for reprocessing.

5.3.4.3 Processing

Processing requirements (MFLOPS) by release and for each DAAC are summarized in Table 5.3-2 Phasing factors and machine efficiency have been taken into account, with machine efficiency assumed to be 25%.

The Technical Baseline establishes the processing load requirements (i.e. MFLOPs) and the data volume requirements (i.e. GB/day) for EOSDIS.

5.3.4.4 Static Analysis of AHWGP Data

A static (spreadsheet) analysis of AHWGP data for Ir1 and Release A has been performed and summarized in Table 5.3-4. The AHWGP data is based on the November 1994 Technical Baseline. ESDIS phasing factors are applied. Additionally, machine efficiency of 25% is accounted for. Values for Release A are presented incrementally. Total capacity values at Release A is the summation of Ir1 capacity and Release A increment. The static analysis estimates peak processing (MFLOPs), peak I/O bandwidth, and disk volume. Further analysis was performed to estimate disk and network bandwidths, but is not considered within the scope of this paper. Host attached disk was assumed for Release A. The I/O bandwidths for LaRC based on static analysis were 1 MB/s at Ir1 and 13 MB/s for Release A. The I/O bandwidth requirements at MSFC was lower than 1. These values were factored up to reflect representative vendor minimum I/O bandwidths.

Table 5.3-4. Static Analysis of AHWGP Data

DAAC	Release	Peak MFLOPs	Peak I/O Bandwidth	Disk Volume
LaRC	Ir1 Rel A	17 +3416	1 MB/s +12 MB/s	22 GB TBD
MSFC	Ir1 Rel A	2 +16	0.07 MB/s No change	186 MB TBD
EDC	Ir1 Rel A	120 +57	16 MB/s +156 MB/s	23 GB TBD
GSFC	Ir1 Rel A	2150 +2040	250 MB/s +1 MB/s	< 75 GB TBD

LaRC and MSFC are operational sites for TRMM CERES and LIS, respectively. At LaRC, 17 MFLOPs are required at launch. However, 1100 MFLOPs are needed 6 months after launch and will be reflected into the processing requirement for LaRC. Therefore, LaRC will require 1100 MFLOPs at Ir1 and an additional 2333 MFLOPs at Release A to satisfy the total Release A requirement of 3433 MFLOPs. EDC provides early Ir1 support for ASTER, and Release A support for MODIS and ASTER. GSFC provides early Ir1 support for MODIS. For GSFC, the MFLOPs at Ir1 assumes 50% of the 0.3 X phasing factor. An example of the processing requirements for LaRC is presented below primarily to provide some insight into the extent of scalability and phased performance requirements.

I/O bandwidth at the CPU provides a good estimate of the amount of data movement coordinated by the CPU. A host attached staging disk is assumed. When a Data Processing Request to the Data Processing System, the data is staged to the staging disk by the CPU. After the completion of processing, output files to be archived are destaged from the staging disk. The CPU coordinates read/write operations during processing. The estimate of average I/O bandwidth (BW) at the CPU is:

(1) I/O BW @ CPU (MB/s) = (2 x Staging Volume + 2 x Destaging Volume + Volume of I/O) x Activations per Day/ 86,400

The I/O bandwidth estimate is based on spread sheet analysis, where the estimate is a time average over one day. Using the AHWGP data in Table 5.3-3, average I/O bandwidth is computed for each PGE and accumulated. A peak I/O bandwidth estimate was made by multiplying the average value by a factor of 5. More accurate peak estimates will be provided by the dynamic model.

Average local disk bandwidth for a Host attached disk is estimated by the relationship:

(2) Disk BW (MB/s) = (Staging Volume + Destaging Volume + Volume of I/O Operations) x Activations per Day/ 86,400

An estimate for average local disk bandwidth supporting TRMM CERES is 0.4 MB/s using spreadsheet analysis. The static analysis results will be refined by the dynamic model.

<u>Example</u>

Using the processing requirements derived from AHWGP data, the rationale for determining processing platforms is shown. This example shows the processing requirements based on AHWGP data for the period 2 years prior to launch (L-2) to 2 years after launch (L+2) with phasing and machine efficiencies taken into account. The LaRC DAAC is used in this example.

- TRMM launch is August, 1997
- Ir1 installation is September, 1995 - November, 1995
- AI&T/I&T is from January, 1996 - January, 1997
- Release A operations run from January, 1996 - October, 1997

Table 5.3-5 Total Processing Load for TRMM/ CERES at LaRC

Processing (MFLOPS) Requirements	Phasing	Epoch	Comments
1100 - 3,433	L-2 < t < L-1: 0.3x (where x = processing load)	12/95 - 12/96	1100 MFLOPs required 6 months after launch for Ir1 Additional 2333 MFLOPs required for Release A
3,433 - 17,157	L-1 < t < L+1: 1.2x	12/96 - 12/98	Scaled up processing through launch (8/97)

It is emphasized that the focus is on Ir1 and Release A with a "Look Ahead" to Release B. The Ir1 processing requirements have been projected from the 17 MFLOPs to 1100 MFLOPs, which is needed 6 months after launch. Release A requirements are 3,433 MFLOPs indicated in the table. Adequate scalability can be achieved by an SMP, which is readily achieved by a representative SMP(s) containing 12 CPUs rated at 300 MFLOPs (peak) each. Additionally, it is anticipated that with improved performance over the next couple of years, technology insertion can further expand scalability.

A migration path for large scalability is a SMP cluster interconnected via a high performance switch (e.g. HIPPI or Fiber Channel), which can support processing requirements from 17 to 17,157 MFLOPs. The representative SMP(s) will be upgraded in several phases:

- minimal configuration SMP-H for Ir1: 4 CPUs @ 300 MFLOPs each to support 1,200 MFLOPs (Peak)
- additional SMP-H for Release A: 8 CPUs @ 300 MFLOPs to support additional 2,400 MFLOPs (Peak). Additional SMP supports "Fail-Soft" environment
- maximum configuration SMP-H: can be scaled to support up to approximately 7,200 MFLOPS (7.2 GFLOPs) for each SMP
- high scale processing via SMP cluster interconnected via high performance switch to support reprocessing requirements (> 7.2 GFLOPs).

5.3.4.5 Data Processing Hardware Provided Capacity

Table 5.3-6 summarizes the AHWGP required capacities vs. the recommended data processing platforms provided capacities. The AHWGP required capacities were derived from Tables 5.3-3 and 5.3-4 which reflected raw AHWGP data and a roll-up summary, respectively. Required capacities were modified in Table 5.3-6 with the modications and assumptions summarized in notes.

Table 5.3-6 AHWGP Required Capacity vs. Provided Capacity

	AHWGP Required Capacity					Provided Capacity		
DAAC	Release	Peak MFLOPs	Peak I/O Bandwidth	Disk Volume	Platform	Peak MFLOPs	Peak I/O Bandwidth	Disk Volume
LaRC	Ir1	1,100	25 MB/sec	30 GB	SMP (4 CPU)	1,200	320 MB/sec	30 GB
	Rel A	+2,333	No Change	No Change	+SMP (8 CPU)	+2,400	+320 MB/sec	No Change
MSFC	Ir1	10	25 MB/sec	5 GB	Uniprocessor WS	125	100 MB/sec	5 GB
	Rel A	No change	No Change	+5 GB	+Uniprocessor WS	+125	+100 MB/sec	+5 GB
EDC	Ir1	120	25 MB/sec	35 GB	SMP (2 CPU)	600	320 MB/sec	35 GB
	Rel A	+57	+156 MB/sec	+35 GB	No Change	No Change	No Change	+35 GB
GSFC	Ir1	2,150	+250 MB/sec	75 GB	SMP (8 CPU)	2,400	640 MB/sec	75 GB
	Rel A	+2,040	No Change	No Change	+6 CPU	+1,800	No Change	No Change

Notes:

(1) ESDIS phasing factors and machine efficiencies are applied.

(2) LaRC required MFLOPs capacities at Ir1 was projected ahead to 6 months after launch (i.e. 1100 MFLOPs)

(3) MSFC MFLOPs per static analysis was less than 1 MFLOPs. 10 MFLOPs was assumed.

(4) GSFC MFLOPs at IR-1 assumes 50% of 0.3 X phasing factor (i.e. 2150 MFLOPs)

(5) I/O bandwidth per static analysis for LaRC and MSFC was 1 MB/s and < 1 MB/s, respectively. Representative vendor minimum was 25 MB/s and was assumed.

(6) LaRC disk volume estimate based on DAAC manager estimate. EDC disk volume estimate based on instrument team input.

(7) Shaded rows for EDC and GSFC sites indicate AI&T environment. LaRC and MSFC are operational sites for Ir1/Release A.

Provided capacities are based on representative vendor inputs. In a high-end SMP, each CPU was assumed to be rated at 300 MFLOPs (peak) with each I/O card rated at 320 MB/s (Peak). The provided capacities were sized to satisfy AHWGP required capacities in Table 5.3-5. The high-end SMP provides very good processing capacity and I/O bandwidth, and scalability and is very cost effective through Release A. In the case of the operational sites (i..e LaRC, MSFC), a second processor is provided for a fail-soft capability.

The candidate platform for LaRC is a high-end SMP scaled to support the required phased performance as illustrated in Table 5.3-5. A 4-CPU configuration SMP has been selected for Ir1. It meets immediate performance requirements and was selected because it is a minimum risk approach, providing a smooth transition from IR-1 to Release A, requiring minimal regression testing. This solution provides reasonably good scalability required for the releases considered. The recommendation for the MSFC DAAC is a uniprocessor workstation or server. A representative uniprocessor supporting 125 MFLOPs can readily support each site for the life of the contract based on the processing requirements. Similarly, a second uniprocessor is provided in Release A to support fail-soft capability. The recommended platform for EDC is a SMP scalable to meet phased performance for AIT/I&T, standard processing and reprocessing. Since EDC only supports AI&T for Release A, a second processor is not provided.

Total EDC processing requirements are 177 MFLOPs for Release A. The processing at EDC is split between the ASTER and MODIS instruments. Considering the scalability, the EDC processor platform is sized for a minimum SMP-H (e.g. 2 CPUs) for Ir1 and Release A. Sizing of SPRHW CIs are derived from static (or spreadsheet) analysis of the Technical Baseline. Peak values have been estimated. More precise estimates of dynamic peak CPU processing performance, I/O bandwidth, and staging values will be determined by system modeling results and applied as soon as they are available. The site specific characteristics of the SPRHW complement are provided in the DAAC unique appendices.

The recommended platform for GSFC is a high-end SMP (SMP-H) configuration (e.g. 8 CPUs) to support MODIS for Ir1 and an additional 6 CPUs for Release A. GSFC supports AI&T in Release A and, therefore, does not require a backup processor. Parallel processors provide excellent processor performance at a relatively low cost. The SMP architecture, utilizing multiple CPUs tightly coupled, can currently provide peak processing performance of 5.4 GFLOPs and a peak I/O bandwidth of 1.28 GB/sec. An 18 CPU SMP configuration is a representative configuration providing this capability. Another representative parallel processor in the Unix cluster class provides a 6.6 GFLOPs (peak) processing performance and 1.2 GB/sec I/O bandwidth that is the most cost effective in this class (i.e. $139/MFLOP). Although processing performance and I/O bandwidth are somewhat less than a vector super computer on a single SMP or Unix cluster computer basis, increased performance can be readily achieved by a clustered computer configuration. These architectures have the added advantage of a 64 bit architecture that allows more addressable space and higher computational accuracy and resolution. Providing a clustered configuration of SMPs can meet the processing performance requirements of CERES, MODIS, and MISR.

5.3.4.6 Scalability

Scalability is the ability to increase processing capacity with minimum impact on both hardware and software. From a hardware standpoint, it is the ability to add on processors; from a software standpoint, it is the ability to provide incrementally better performance with minimum tuning. This ability to expand processing with minimum effect on existing operations can be achieved in several ways, including but not limited to:

- adding new processing strings
- adding CPUs to existing host computers
- adding computer systems to existing strings
- adding subnetworks to support additional inter and/or intra-string I/O and communications
- technology refresh

The ability to select a processor class which satisfies increasing performance requirements in incremental releases and a spectrum of scientific missions including TRMM, AM-1, and PM-1 is one of the major challenges of the design. Initially, processor performance requirements must satisfy the TRMM mission with CERES and LIS instruments in Release A.

An SMP configuration provides reasonable scalability (up to 32 processors) and is more than adequate for most sites included in Release A. For Release B, the scalability can be extended (primarily at LaRC and GSFC) by employing clusters of SMPs within a high-speed network.

Total required processing capacities (MFLOPS) for each DAAC are summarized in Figure 5.3-2. This figure represents the theoretical peak MFLOP requirements for each DAAC using a logarithmic scale for releases IR-1, A, and B. It represents total mission processing

requirements including MFLOPs required by TRMM and AM-1 AI&T. The purpose of presenting this figure is to illustrate the scalability required from IR-1 to Release A, and to Release B.

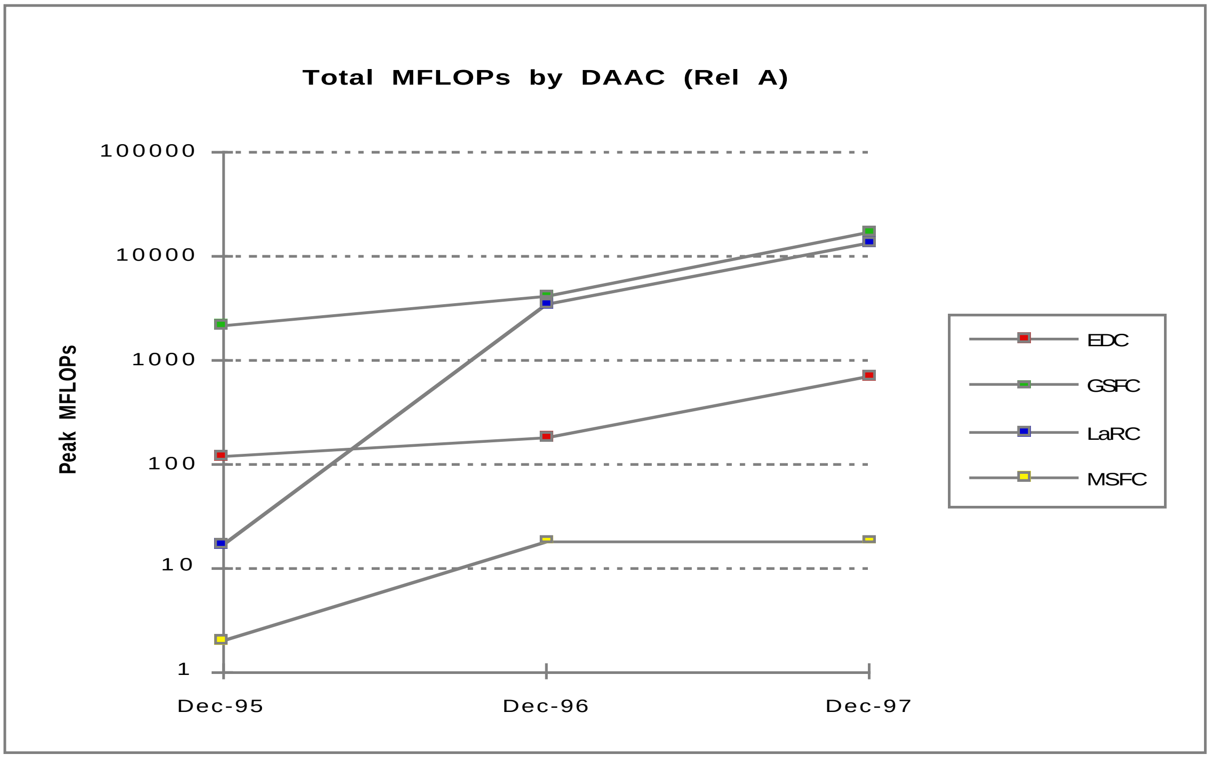

Figure 5.3-2 Total MFLOPs by DAAC

Scalability requirements are most significant with LaRC and GSFC. LaRC processing requirements vary from 17 MFLOPs in Ir1 to a total of 3,433 MFLOPs in Release A and a total of 13,353 MFLOPs in Release B. Referencing Table 5.3-1, Release A processing requirements for LaRC can be met by a high-end SMP with adequate margin. Release B would require a cluster of at least two maximum configuration high-end SMPs for LaRC.

GSFC processing requirements vary from 2,150 MFLOPs in Ir1 to a total of 4,190 MFLOPs in Release A and a total of 16,762 MFLOPs in Release B. Similarly, GSFC processing requirements for Ir1 and Release A can readily be satisfied by a high-end SMP. Release B requirements suggest a cluster of at least three high-end SMPs connected by a high speed switch or fibre channel.

According to Table 5.3-6, I/O bandwidth requirements for Ir1 and Release A for LaRC, MSFC, and EDC are relatively low (< 25 MB/s) and can readily be satisfied. The GSFC I/O bandwidth requirement of 251 MB/s can be met by a representative high-end SMP. A representative high-end SMP is scalable to over 1 GB/s (Peak). I/O bandwidth requirements for MODIS in later releases are anticipated to be very high. Further analysis is required utilizing the system model to assess peak and sustained I/O bandwidths.

Scalability provides the ability to add additional CPUs and linearly increase processing performance. One representative vendor has indicated a scalability of 0.9 for the addition of each CPU up to 6 CPUs (i.e. 3.5 GFLOPs). A second representative vendor indicates scalability up to an 18 CPU configuration corresponding to a processing performance of 5.4 GFLOPs. Prototyping results indicated a more limited scalability with the knee of the curve at approximately 12 CPUs. Scalability associated with vector super computers is somewhat limited, nominally in the range of 32:1. MPPs offer the best potential for high scalability. However, as with SMPs, the MPPs resulting performance can be affected by the degree of parallelization of the science software and tuning.

5.4 Advantages and Disadvantages

The major advantages and disadvantages of each candidate class are summarized in Table 5.4-1. This table primarily compares cost, performance, and scalability parameters.

The major criteria for candidate platform class(es) selection are cost, performance, and scalability and evolvability.

The uniprocessor workstation platform class is attractive for lower performance requirements primarily due to low list price as well as low $$/MFLOPs (i.e. 205$/MFLOPs (average). It is ideal for such applications as science processors at MSFC and for Processing Queuing Management and Planning workstations.

A workstation farm (i.e. cluster of workstations) connected via a LAN provides excellent scalability. Further, it has the advantage of utilizing low cost commodity items in a network. Limitations are posed by low network bandwidth and high latency. An extension of the workstation farm is a SMP cluster. A high performance network (e.g. HIPPI) or switch (e.g. fibre channel) significantly improves interconnect bandwidths.

A vector supercomputer provides high processing performance and I/O bandwidth with good scalability. It is very efficient in vector computation. However, it has a relatively high list price and high $$/MFLOPs (i.e. 751$/MFLOPs (average). This platform class will be reconsidered for Release B and beyond.

Table 5.4-1 Platform Classes Compared

Platform Class	Advantages	Disadvantages
Uniprocessors	• Low cost commodity item • Flexible • Controlled by user	• Very limited scalability • Limited performance
Workstation Farm	• Low Cost • Flexible • Scalable • Provides special purpose parallel environment	• Performance based on interconnectivity, load balancing, network bandwidth • Less predictable environment (total performance must take into account individual user profiles) • Programming a heterogeneous cluster can be complex • Loose coupling desired
Vector Super computer	• High performance • High efficiency / less parallelization effort	• Higher Cost
SMP-H	• Low $$/MFLOPs cost • Good scalability • High performance • Graceful degradation	• Scalability limited to 64 processors
SMP-L	• Low $$/MFLOPs cost • Moderate scalability • Graceful degradation	• Scalability limited to 8 processors
MPP	• Highly scalable • High RMA • Graceful degradation • Load balancing • High I/O bandwidth • High performance, with tuning	• Programming effort required to tune to peak efficiency • However with parallelization tools, portability is made easier • Expensive

The Symmetric Multiprocessors (SMPs) provide excellent cost/performance, very good scalability, graceful degradation, and releatively low risk. The average $$/MFLOPs for a low- and high-end SMP are 147 and 176 $/MFLOPs, respectively (Reference Table 5.3-1). The low-end SMP is typically scalable from 1 to 8 CPUs. The high-end SMPs can be scalable from 2 to 64 CPUs. and approximately up to 6 GFLOPs (Peak). Interconnect bandwidth within SMPs are very high. SMPs allow multiprocessing capabilities in addition to supporting conventional sequential software. SMPs are equivalent to multiple workstations in a box with identical processors and a single operating system.

Massively Parallel Processors (MPPs) provide high performance and highly scalable (greater than 64 CPUs) and potentially high I/O bandwidth. There is programming effort to tune to a high efficiency and thus can be more difficult to port existing applications. The average $$/MFLOPs (i.e. 319 $$/MFLOPs) are significantly higher than the SMP. This platform class will be reconsidered for Release B and beyond as the algorithm parallelization techniques and automatic tuning become more mature.

5.5 Risk

Two kinds of risk are considered for in this evaluation:

- technical
- schedule

Business risk is an important component of this category, and will be taken into account in the procurement cycle.

5.5.1 Technical Risk

The major elements of risk associated with the selection of the hardware processor includes the following: (1) Confidence level of processing requirements, (2) Benchmarking science software, and (3) Transitioning from 32 bit to 64 bit architecture.

Confidence level of processing requirements can influence the selection of hardware platforms. Requirements for AIT are estimated and requirements for QA and monitoring are just beginning to evolve. There is still a level of uncertainty attached to these figures. Additionally, science algorithms will be refined and updated every six months during the first year after launch. Another uncertainty is the ratio of standard products vs ad hoc processing in the future. Consequently, this potential risk suggests emphasis on scalability and expandability and at the same time staying within cost constraints. Since processor costs are estimated to drop off 21% each year, it would suggest that cost should be minimized early or even delayed and to provide a design that is highly flexible and expandable. Installing a vector super computer early is expensive and does not have enough scalability and expandability. A loosely coupled workstation farm runs the risk of inadequate communications bandwidth. The SMPs or a Unix cluster offer a relatively low risk approach. The SMP architecture is conceptually simple, provides a low cost to performance ratio, is portable, code development is easier, and familiarity has been provided by prototyping experience. A major cost advantage is that these machines use commodity chips. Clustering SMPs or a Unix cluster can provide the scalability and expandability. NASA Ames has demonstrated a 64 bit super computing cluster comparing how different architecture's execute parallel jobs, The Ames installation utilizes heterogeneous clusters.

Processor Efficiency: Prototyping based upon a heritage ECS algorithm (SeaWinds) indicated the following efficiencies for one processor: 20% for SGI Challenge, 17% for Cray T3D, and 11% IBM SP2 (Reference 430-TP-006-001). It should be noted that efficiencies are dependent upon hardware configuration as well as the algorithm. Vendors provide industry standard benchmarks (e.g. Specint92, Specfp92, Linpack 1000), that provide some indication of performance, but are not necessarily representative of the actual science algorithms.

Transitioning from 32-bit architecture to 64-bit architecture: Most science processing platforms at the DAACs will be of 64-bit architecture to accommodate greater scalability for Release A and beyond. However, science software delivered by the Instrument Teams (ITs) on 32-bit platforms could pose a risk in terms of software portability. The portability issues for ECS science software integration was presented at the Science Software Integration & Test Workshop (Reference 726-PP-002-001). To temporarily mitigate the risks due to software portability, compiler features available on the processing platforms will be used to maintain the 32-bit environment on a 64-bit platform. Using such compiler features can degrade overall performance. This should be considered only as a temporary measure to aid science software integration for IR-1. We expect science software developers, at some point, to develop software on 64-bit platforms so that portability problems can be minimized.

5.5.2 Schedule Risk

In evolution of science software immediately following launch of an instrument could pose a risk in terms of schedule. Also, differences in byte-ordering of the architecture can translate into a schedule risk

5.6 Suitability

Suitability includes additional considerations to be examined along with the other quantitative selection criteria in order to determine the final platform recommendations.

- Algorithm Parallelization
- Adaptability of Science Software
- Phased Performance Requirments
- Architecture Design impacts

5.6.1 Algorithm Parallelization

The major consideration in selecting the processor class is the ability to run parallel programs especially when large processing requirements are called for in Release B. An SMP provides the ability to parallelize programs in shared memory (the easiest paradigm for parallel programs) and distributed memory modes. A variety of tools are available to parallelize science software. The ECS Science and Technology Lab (STL) prototyping (Reference: 440-TP-008-001, 194-00569TPW) of science processing has demonstrated parallel program development based primarily on parallelization tools.

5.6.2 Adaptability of Science Software

The major consideration in selecting the processor class is that science software parallelization must be accounted for in future systems. Processor efficiency can significantly affect the performance of science software. Prototyping work at the ECS STL provide information on processor efficiency based on real science software (Ref:430-TP-006-001). Efficiencies were calculated for symmetric multiprocessors, workstation clusters, and massively parallel processors. Performance measurements for both single and multiprocessing paradigms were evaluated. Although the processor efficiency measurements were based on a single science software (SeaWinds), nevertheless it provided valuable information for systems performance modeling. A processor efficiency of 25% appears reasonable for sequential processing for this class of machines currently considered for Ir1 and Release A.   If the algorithms are highly parallelized with CPU intensive processing, there is a significant benefit in parallel processing. It is important to recognize that industry standards (e.g. Linpack 1000, Specint92, Specfp92) can only provide a rough indication of CPU performance and scalability, since the industry standards do not necessarily characterize the science software implementation. The ultimate test is how the processors will perform with the actual science software used as the benchmark. The need for parallelization to effectively utilize a given processor configuration must be balanced against the need for portability, as required to preserve options for future hardware evolution.

5.6.3 Phased Performance Requirements

The ability to select a processor class satisfying increasing performance requirements in incremental releases and satisfy a spectrum of scientific missions including TRMM, AM-1, and PM-1 is one of the major challenges of this trade study. Initially processor performance requirements must satisfy the TRMM mission with CERES and LIS instruments in release A. In a subsequent releases, requirements imposed by algorithms that evolve into "tall poles" (e.g. MODIS, MISR) must be satisfied. The ability to linearly increase processor performance incrementally defines scalability. This allows the capability to incrementally add CPUs and I/O channels resulting in a near linear increase of MFLOPs and I/O bandwidth performance in a cost effective manner with minimum program risk.

5.6.4 Architecture Design Impacts

The third factor which plays a major role in selection of the processor class is the ability to consider the architectural design impacts. For example, if a workstation cluster were selected, it provides a low cost solution (i.e. $/MFLOPs) for each hardware platform. However, if the workstations are interconnected on a LAN, the interconnect bandwidths are three orders of magnitude slower than a memory to memory interconnect in a tightly coupled cluster of SMPs. Another example of architecture design impact involves how well the recommended platform fits in with the existing platforms. Considerations of this nature include converting to or integration of different operating systems, code conversions, 32-bit vs. 64-bit, etc.

5.7 Evolvability

Evolvability is the ability to accommodate an incremental development and allow technology insertion. It also involves the ability to expand gracefully, with minimal throwaway. Evolvability allows the selection to initially satisfy requirements for IR-1, then phase into release A AI&T and I&T, standard processing, and then reprocessing, A similar process would follow for subsequent releases. SMP clusters or UNIX clusters provide excellent evolvability characteristics. NASA Ames has successfully demonstrated integration of heterogeneous clusters, which would lead to a lower cost approach that would support evolvability. Evolvability for MPPs is highly dependent on the parallelization of algorithms and the development of parallelization compilers.

5.8 Reliability/Maintainability/Availability (RMA)

SMPs inherently provide a significantly higher reliability and availability for the same hardware since there is shared memory and multiple paths between the major elements. Additionally, SMPs provide a more graceful degraded mode capability. All other configurations must provide additional redundancy to achieve the same reliability and availability. Thusfar, vendor data has been very limited in this area. Therefore, quantitative comparisons are not available at this time.

5.9 Portability

It is anticipated that open standards, ANSI standards, and C++ and F77 will be met by all vendors. Prototyping has provided additional familiarity and experience with the operating systems and other portability issues.

6. Selection Criteria and Evaluation

6.1 Selection Criteria

This section identifies the selection criteria to be used in the evaluation phase. Each of these criteria are weighted in terms of relative importance to the project on a scale from 1 to 10 where 10 is the most important. Components of suitability are critical, thus merit a "pass/fail" scoring.

The purpose of this section is to try to quantify the relative importance of each of the major selection criteria and provide a figure of merit.

- cost
- performance
- risk
- reliability/maintainability/availability (RMA)
- scalability
- evolvability
- suitability (other factors which need to be taken into consideration include the extent that science software can adapt to a production environment without losing CPU efficiency, throughput, satisfy phased performance requirements and have minimal architecture design impacts.)

The actual selected weights will be based on an average of surveyed results obtained from NASA, the ECS design team and investigative teams. In the meantime, initial weights have been assigned and are summarized below. These weight factors will be adjusted over time to reflect the survey results.

Table 6.1-1. Weighted Selection Criteria

Criteria	Components	Criteria Weight
Cost	Development Effort	10
	Life cycle cost (acquisition cost, maintenance cost, support cost) (normalized to $/MFLOPS)	
Performance	Using MFLOP Performance now, next iteration will use actual Science software Benchmark Performance	9
	I/O Bandwidth	
	Working Storage	
Risk	Technical Risk	9
	Schedule Risk	
	Business Risk	
RMA	Reliability	8
	Maintainability	
	Availability	
	Degraded Modes	
	Recovery	
Scalability		9
Evolvability		8
Suitability	extent that science software can adapt to a production environment without losing CPU efficiency (adaptability of science software), throughput, ability to satisfy phased performance requirements and have minimal architecture design impacts	Pass/Fail

6.2 Figure of Merit Evaluations

Based on a questionnaire sent out to vendors, list price, performance capabilities, and other technical specifications have been obtained for several representative platforms and will be used for this initial evaluation. The vendor data was averaged for each category. The primary evaluation criteria are cost, performance, risk, and scalability.

The Figure of Merit is defined as the product of the weighting factor (Reference Table 6.1-1) and scoring for each evaluation criteria., where 10 is the highest score. The score is indicated within the bracket in Table 6.2-1. For example, the weight for cost is 10 and the score for SMP-H is 9. Therefore, the Figure of Merit is 90 for cost in this instance.

It should be recognized that this Figure of Merit can be subjective and variable. It is dependent on the weights selected and score for each criteria. Sensitivity analysis is recommended to understand the effects of changing weighting or scoring. However, it is expected that these scoring refinements will not drastically change the recommended platforms.

Risk, RMA, evolvability, and portability will be rated more subjectively using the characteristics exhibited by each platform category. Suitability is rated based on processing specific to each DAAC.

Table 6.2-1 Platform Class Evaluation: Summary

Platform Class	Cost	Performance	Risk	Scalability	Evolvability	Total
Uniprocessor	95 (9.5)	45(5)	63 (7)	45 (5)	40 (5)	**288**
Workstation Farm	85 (8.5)	63 (7)	72 (8)	81 (9)	64 8)	**365**
Vector Supercomputer	*50 (5)*	*85.5 (9.5)*	*72 (8)*	*72 (8)*	*72 (7)*	***351***
SMP-H	*90 (9)*	*81 (9)*	*81 (9)*	*63 (7)*	*64 (8)*	***379***
SMP-L	*100 (10)*	*63 (7)*	*81 (9)*	*54 (6)*	*64 (8)*	***362***
MPP	60 (6)	85.5 (9.5)	54 (6)	81(9)	72 (9)	**353**

The SMP-H has the highest FOM score, and is therefore recommended for LaRC, GSFC, and EDC for Release A. Cost is a major criteria in selecting SMP-H. In the case of GSFC and LaRC, the only real choices of processor platforms from a performance viewpoint for release A are the SMP-H, vector supercomputer, and MPP (See Tables-5.3-1 and 5.3-2). The SMP-H's average normalized cost is 176 $$/MFLOP as compared to 751 $$/MFLOPs for the vector supercomputer and 300 $$/MFLOPs for the MPP. Average list price is less than 40% of the cost the vector super computer, which is its nearest competitor.

The SMP-H is relatively low risk. It can handle sequential processing and can transition to parallel processing. It is simpler to program and has been demonstrating during prototyping.

7. Recommendation/Conclusion

The recommended processor platform classes for Ir1 and Release A are the high-end SMP (SMP-H) for LaRC, EDC, and GSFC and the uniprocessor workstation for MSFC. These processors are contained within the Science Processing HWCI at each site. A uniprocessor workstation is recommended for the Processing Queuing Management function at all sites. This recommendation is based upon the November 1994 PDR Technical Baseline containing AHWGP capacities, scalability of hardware, evolvability of hardware, and cost performance. For the operational sites (LaRC and MSFC), a minimum configuration. SMP-H is utilized to support AI&T. A second SMP-H is added in Release A to support the fail-soft requirement. A summary of the platform classes recommended for Ir1 and Release A is contained in Table 2.1-1.

The SMP platform class was selected for the following reasons:

- Best cost/performance with more capacity for the money
- Environment supports conventional sequential in addition to symmetric anddistributed memory parallel processing, thereby reducing risk
- Interconnect bandwidths between processors are much higher than external bandwidths
- Results of ECS STL prototyping

The focus is on Ir1 and Release A with a "Look Ahead" to Release B for scalability and evolvability. Although a SMP platform class is selected for Ir1 and Release A, the other platform classes (i.e. Vector supercomputer, MPP) will be readdressed for Release B and beyond.

This study is being conducted in a minimum of two phases: (1) Static analysis and (2) Dynamic analysis using ECS System Performance Model. In the first phase, performance analysis was based on the static model (i.e spread sheet) and produced time-average results. Since there is adequate performance margin for Ir1 and Release A for all sites, the platform classes recommendations will be retained. The second phase of the trade study incorporating the ECS Systems Performance Model will provide both peak and sustained performance and will provide better simulation fidelity.

Although this study is slated for completion at PDR, there are a number of reasons why it should be re-addressed on an as-needed basis through Release A CDR as a minimum:

- Further AHWGP activities will result in new analyses/ technical baselines, resulting in refined static and dynamic analysis.
- Completion of the system level dynamic model will permit increased analysis fidelity and exploration of topology and technology "what if" analysis.
- Competitive procurements are phased over time for the releases.
- Continued work with the science software teams will provide additional derived requirements to be considered for inclusion within the Data Processing Subsystem.

Abbreviations and Acronyms

AI&T	Algorithm Integration & Test
AITHW	Algorithm Integration & Test Hardware
AHWGP	Ad Hoc Working Group on Production
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BONeS	Block Oriented Network Simulation
CDR	Critical Design Review
CI	Configuration Item
CPU	Central Processing Unit
DAAC	Distributed Active Archive Center
DCE	Distributed Computing Environment
DEC	Digital Equipment Corp.
DID	Data Item Description
DMP	Distributed Memory Processing
ECS	EOSDIS Core System
EDC	EROS Data Center
EOSDIS	EOS Data and Information System
ESDIS	Earth Science Data and Information System
FOM	Figure of Merit
GB	Giga-bytes
GSFC	Goddard Space Flight Center
HIPPI	High Performance Parallel Interface
HWCI	Hardware Configuration Item
IBM	International Business Machines Corp.
I/O	Input/Output
IR-1	Interim Release- 1
I&T	Integration & Test
LAN	Local Area Network
LaRC	Langley Research Center

LIS	Lightning Imaging Sensor
MB	Mega-bytes
MFLOPs	Millions of Floating Point Operations per Second
MIPs	Millions of Instructions per Second
MISR	Multi-Angle Imaging Spectro Radiometer
MODIS	Moderate Resolution Imaging Spectrometer
MPP	Massively Parallel Processor
MSFC	Marshall Space Flight Center
PDR	Preliminary Design Review
QA	Quality Assessment
RISC	Reduced Instruction Set Computing
RMA	Reliability/ Maintainability/ Availability
SCF	Science Computing Facility
SDPS	Science Data Processing Segment
SGI	Silicon Graphics Inc.
SMP	Symmetric Multiprocessor
SMP-L	Symmetric Multiprocessor- Low end
SMP-H	Symmetric Multiprocessor - High end
SPRHW	Science Processing Hardware
STL	Science and Technology Laboratory
RAID	Redundant Arrays of Inexpensive Disks
TRMM	Tropical Rainfall Measuring Mission
$$/MFLOPs	Dollars per Millions of Floating Point Operations per Second